The purpose of rating quality of evidence differs in an overview, as compared to guidelines or recommendations

Pollock, Alex; Brady, Marian C.; Farmer, Sybil E.; Langhorne, Peter; Mead, Gillian E.; Mehrholz, Jan; Wiffen, Philip J; van Wijck, Frederike

Published in:
Journal of Clinical Epidemiology

DOI:
10.1016/j.jclinepi.2016.01.001

Publication date:
2016

Document Version
Author accepted manuscript

The purpose of rating quality of evidence differs in an overview, as compared to guidelines or recommendations

Response to letters:

Gionfriddo MR. Subjectivity is a strength: A comment on "An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews"

Murad MH, Mustafa R, Morgan R, Sultan S, Ytter-Y, Dahm P. Rating the quality of evidence is by necessity a matter of judgement.

Authors:

Alex Pollock[a], Marian C Brady[a], Sybil E Farmer[a], Peter Langhorne[b], Gillian E Mead[c], Jan Mehrholz[d], Philip J Wiffen[e], Frederike van Wijck[f]

[a]Nursing, Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Glasgow, UK

[b]Academic Section of Geriatric Medicine, University of Glasgow, Glasgow, UK

[c]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

[d]Wissenschaftliches Institut, Private Europäische Medizinische Akademie der Klinik Bavaria in Kreischa GmbH, Kreischa, Germany

[e]Pain Research and Nuffield Department of Clinical Neurosciences (Nuffield Division of Anaesthetics), University of Oxford, Oxford, UK

[f]Institute for Applied Health Research and the School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, UK

Corresponding author:
Alex Pollock,
Nursing, Midwifery and Allied Health Professions Research Unit
Glasgow Caledonian University
Buchanan House
Cowcaddens Road
Glasgow G4 0BA
UK

Tel: +44 (0)141 331 8100

Email: alex.pollock@gcu.ac.uk

The purpose of rating quality of evidence differs in an overview, as compared to guidelines or recommendations.

We welcome the points raised within the letters by Gionfriddo (2015) and Murad et al (2015) in response to our paper and are pleased that our paper has prompted a wider debate. The matters raised are not new to us, and we spent considerable time reflecting on and debating similar issues prior to our decision to adopt the algorithmic approach. We do not believe that our views are discordant with those of Gionfriddo (2015) and Murad et al (2015), as we do not consider that there is a simple dichotomous decision in which one can be either in favour of or against the use of objective criteria as part of the process of assessing quality of evidence. However, in light of the points raised, we take this opportunity to advocate caution to those considering use of our proposed approach for other purposes, and to repeat the statement within the original paper that our algorithm is specific to the body of evidence relating to interventions to improve upper limb function after stroke. Our paper does not promote the use of our specific algorithm *per se*, but rather purports that the approach may be applicable to other overviews or evidence syntheses.

Gionfriddo (2015) highlights that the rating of an overall body of evidence may depend on the purpose for which the evidence is assessed. We agree. We had a particular purpose and reason for rating the quality of evidence which justifies and supports our algorithmic approach. The principal purpose of an overview of reviews is "to compile evidence from multiple systematic reviews of interventions into one accessible and usable document", providing clinical decision makers with an "exhaustive list of reviews relevant to a specific decision" (Becker 2011). Our overview aimed to signpost clinical decision makers to the systematic review(s) we judged to contain the highest quality evidence relating to specific interventions; the purpose was *not* to make treatment recommendations, but rather to direct people to best evidence. We acknowledge that for a different purpose, such as formulating clinical guidelines, our algorithmic approach may lack the essential subjectivity required to support interpretation and application of the evidence to a particular context. We are in accordance with the fact that "evidence alone is not enough" (Gionfriddo 2015), and this is demonstrated throughout the Cochrane overview where we clearly stress that "clinical application of evidence will depend on specific details of an individual patient or setting, or both, and clinical decisions will require expert clinical reasoning and judgement if available evidence is to be interpreted and applied effectively" (Pollock 2014).

Murad et al (2015) rightly clarify that GRADE has been shown to lead to reproducible results when applied by independent assessors with sufficient competency. However we believe it was not lack of competency in GRADE, but rather the volume of comparisons to be assessed, which led our independent reviewers to become aware that their decisions were being heavily influenced by a pattern emerging from objective data. As we report in our paper, during initial attempts to subjectively GRADE the quality of the 127 comparisons included in our overview, both independent reviewers reported using specific data, such as the number of trials and included participants, to inform their decision, and it was this observation which led us to develop our algorithm. Remember that we were rating our confidence in single comparisons (meta-analyses), and not attempting to interpret or rate a body of evidence comprising a number of different (but related) comparisons, as one may do in relation to a guideline recommendation. This brings us back again to the argument of the unique *purpose* of our overview, and when we weighed up the advantages (reproducibility,

transparency, efficiency) and disadvantages (limited opportunities for subjectivity, inconsistency), persuaded us that in this specific situation the advantages outweighed the disadvantages.

Murad et al (2015) argue that "the relative importance of the various domains of quality of evidence varies depending on the situation". This reflects our position, and is why we clearly stated that our algorithm "was developed specifically for our Cochrane overview (Pollock 2014), with careful consideration of the type of evidence included within this overview". We were careful to stress that it was our *approach* which may have relevance, and we did not advocate direct application of our specific algorithm to the assessment of quality of evidence within other overviews.

Where we do disagree with Murad et al (2015) is in relation to our decision to include the quality of the systematic review as a factor within our algorithm. Murad et al (2015) purport that "non-credible" systematic reviews ought to be excluded from overviews of reviews. Within our overview we made careful, and clearly reported, judgements where we identified overlapping overviews, in order that we could include the most up-to-date, comprehensive, methodologically rigorous reviews, and exclude the others. In some instances we identified as many as 10 published systematic reviews synthesising similar evidence relating to the same intervention. A key benefit of our overview is that it signposts clinical decision makers to the 'best' systematic review relating to each single intervention, thus steering clinical decision makers away from the least credible reviews, and signposting best available evidence. Yet once more, any apparent differences in opinion between our group and the views of Gionfriddo (2015) and Murad et al (2015) relate to clarity of *purpose* of rating of evidence. For the purposes of a clinical guideline, excluding reviews which do not meet certain quality criteria may be the correct approach; but within an overview, where the primary purpose is to compile evidence from relevant systematic reviews, the process of signposting the best available systematic review evidence is of central importance.

Thus we do not believe that our approach, as applied to rating the level of evidence within an overview of reviews, is at odds with the views posed by either Gionfriddo (2015) or Murad et al (2015) relating to the necessity for judgement. However the points raised do make us acknowledge the inherent risks of publishing our approach in this paper, many of which relate to risks of misinterpretation or misapplication of our proposed approach. We did not promote our approach as a quick or easy alternative to the application of the GRADE process for all purposes nor do we propose our specific algorithm as one that can be directly applied to other bodies of evidence. Further it is important to highlight that we went through a number of different stages in the development and refinement of our algorithm, involving in-depth discussions during which expert opinions were explored. The time and resources we input into this process promoted efficiency within our project as we planned to rate the quality of evidence of 127 comparisons. With fewer comparisons to rate arguably the relative efficiency would reduce, and potentially shift the balance between advantages and disadvantages.

We postulate that what has prompted much of the debate from both Gionfriddo (2015) and Murad et al (2015) is the extent to which the purpose of rating evidence differs in an overview, as compared to guidelines or recommendations. While Cochrane recommends use of GRADE to rate quality of evidence within overviews, and while our algorithm built on our understanding of the GRADE approach, perhaps our algorithmic approach has moved so far from GRADE that it can no longer be labelled as such. However, regardless of name, our methodological approach has potentially got

implications for assessment of quality of evidence within future overviews, with advantages relating to efficiency, reproducibility and transparency. We look forward to opportunities to debate this further and to work in collaboration with members of the GRADE working group to explore the application of our ideas further. Following discussions with members of the GRADE working group after presentation of this work at the Cochrane Colloquium (Vienna, October 2015), we are in the process of submitting a proposal for a GRADE project group to explore these issues further.

Acknowledgements

This Cochrane overview was supported by a project grant (CZH/4/854) from the Chief Scientist Office (CSO), part of the Scottish Government Health and Social Care Directorate.

References

Becker LA, Oxman AD. Chapter 22: Overviews of reviews. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, van Wijck F. Interventions for improving upper limb function after stroke. Cochrane Database of Systematic Reviews 2014, Issue 11. Art. No.: CD010820. DOI: 10.1002/14651858.CD010820.pub2.